

Environ Model Assess (2010) 15:175–188
 DOI 10.1007/s10666-009-9200-z

Uncertainty from Model Calibration: Applying a New Method to Transport Energy Demand Modelling

Bas van Ruijven · Jeroen P. van der Sluijs ·
 Detlef P. van Vuuren · Peter Janssen ·
 Peter S. C. Heuberger · Bert de Vries

Received: 21 October 2008 / Accepted: 22 July 2009 / Published online: 27 August 2009
 © The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Uncertainties in energy demand modelling originate from both limited understanding of the real-world system and a lack of data for model development, calibration and validation. These uncertainties allow for the development of different models, but also leave room for different calibrations of a single model. Here, an automated model calibration procedure was developed and tested for transport sector energy use modelling in the TIMER 2.0 global energy model. This model describes energy use on the basis of activity levels, structural change and autonomous and price-induced energy efficiency improvements. We found that the model could reasonably reproduce historic data under

different sets of parameter values, leading to different projections of future energy demand levels. Projected energy use for 2030 shows a range of 44–95% around the best-fit projection. Two different model interpretations of the past can generally be distinguished: (1) high useful energy intensity and major energy efficiency improvements or (2) low useful energy intensity and little efficiency improvement. Generally, the first lead to higher future energy demand levels than the second, but model and insights do not provide decisive arguments to attribute a higher likelihood to one of the alternatives.

Keywords Model calibration · Uncertainty · Global energy model · Transport energy use

B. van Ruijven (✉) · D. P. van Vuuren · P. Janssen · B. de Vries
 Netherlands Environment Assessment Agency (PBL),
 P.O. Box 1, 3720 BA Bilthoven, The Netherlands
 e-mail: bas.vanruijven@pbl.nl

D. P. van Vuuren
 e-mail: detlef.vanvuuren@pbl.nl

P. Janssen
 e-mail: peter.janssen@pbl.nl

B. de Vries
 e-mail: bert.devries@pbl.nl

J. P. van der Sluijs · B. de Vries
 Department of Science, Technology and Society,
 Copernicus Institute for Sustainable Development and Innovation,
 Utrecht University,
 Heidelberglaan 2,
 3584 CS Utrecht, The Netherlands

J. P. van der Sluijs
 e-mail: j.p.vandersluijs@uu.nl

P. S. C. Heuberger
 Delft Center for Systems and Control, Delft University,
 Mekelweg 2,
 2628 CD Delft, The Netherlands
 e-mail: P.S.C.Heuberger@tudelft.nl

Abbreviations

AEEI	autonomous energy efficiency improvement
GDP	gross domestic product
GLUE	generalised likelihood uncertainty estimation
IEA	International Energy Agency
IMAGE	integrated model to assess the global environment
IPCC	Intergovernmental Panel on Climate Change
NRMSE	normalised root mean square error
OECD	Organisation for Economic Co-operation and Development
OECD-EO	OECD Environmental Outlook
PEST	parameter estimation
PIEEI	price-induced energy efficiency improvement
SRES	IPCC Special Report on Emission Scenarios
TIMER	The IMAGE Energy Regional Model
UE	useful energy
UEI	useful energy intensity
UNEP	United Nations Environmental Program
WDI	world development indicators

1 Introduction

Uncertainties play a key role in projecting future developments of the energy system. At least two factors contribute to this: (1) the energy system is determined by complex interactions of a wide range of drivers and (2) there is a lack of empirical data. Factors that influence future energy demand and supply include economic activity, developments in economic structure, lifestyle changes and technology development. Our understanding of the interaction of these factors is still limited (and they may range over a wide range of possible outcomes). On top of this, the lack of empirical data complicates the development and calibration of models, especially for developing regions.

Despite limitations in both theory and data availability, a wide range of models has been developed to explore trends at global, regional and national scales. These models are partly developed from different scientific paradigms, which may lead to different interpretations of the past and different expectations of the future [32, 48]. A classic example is the difference between models from a macro-economic tradition (top-down) and those from a technological tradition (bottom-up). These two traditions tend to interpret the present situation differently with respect to energy efficiency ('improvement of energy efficiency leads to higher costs' vis-à-vis 'major opportunities for improvement without substantial costs') and as a result also expect different mitigation costs in the future [23]. Even within one model, however, often different options exist on how to interpret the current and past situation. For instance, macro-economic demand functions often include both income elasticity and price elasticity, which are hard to identify unambiguously in historic data. A different interpretation of the past may lead to different calibrations of the model and uncertainty in future projections. So far, different methods have been used to explore uncertainty in global energy models [12, 30, 53, 65], but relatively little attention has been given to the influence of model calibration on future projections.

The issue of multiple model calibration is closely related to the concept of equifinality, which focuses attention 'on the fact that there are many acceptable representations that cannot easily be rejected and should be considered in assessing the uncertainty associated with predictions' [7]. These 'acceptable representations' are called *behavioural*. The "acceptance criterion" can be defined strictly quantitative (e.g. above a threshold value of a likelihood measure) or more qualitative (e.g. reproduction of trends). At present, calibration of energy models is often done on the basis of the modeller's expert knowledge to identify a single set of plausible parameter values. However, if multiple sets of parameter values are tenable and model projections are sensitive to the parameter values chosen, this practise is questionable [17].

In this context, we have developed a method to automatically calibrate models and obtain sets of parameter values that perform reasonably against historic data. These calibrated sets are obtained by varying the main model parameters within a limited range, choosing an initial estimate in this range and searching consecutively for a (local) optimum to minimise the error between observations and model results. Repeating this procedure many times, initialised at different locations in the parameter space, generates a series of (different) calibrated sets of parameter values. This method is related to both nonlinear regression methods like parameter estimation (PEST) [16] or UCODE [43] and (sequential) Monte-Carlo-based approaches like generalised likelihood uncertainty estimation (GLUE) [8] or SimLab [49].

We apply this method to the energy demand module of the global energy model The IMAGE Energy Regional Model (TIMER) 2.0, a system dynamics model that simulates developments in global energy supply and demand [14, 67]. The TIMER 2.0 model is the energy sub-model of the integrated model to assess the global environment, IMAGE 2.4, that describes the main aspects of global environmental change [10]. In recent years, this model has been used in several global scenario studies like the IPCC Special Report on Emission Scenarios (SRES) [25], the Millennium Ecosystem Assessment [33], UNEP Global Environmental Outlook [54] and the Organisation for Economic Co-operation and Development (OECD) Environmental Outlook [40].

Based on above considerations, the main research question that this paper focuses on is whether in the TIMER model equifinality in calibration can be observed, and if so, what it means for future projections. It should be noted that several uncertainty studies have been performed with the TIMER model [55, 61, 64]. These analyses accepted the model's initial calibration and focused on the spread in model outcomes based on variation in central input values. Moreover, they (and for that matter the same applied to other global energy models) focused on the global level, neglecting interesting underlying trends in different regions. Recent analysis of TIMER found that uncertainty in energy demand trends is a major source of model uncertainty [65]. Therefore, we focus this analysis on the TIMER energy demand sub-model. Within energy demand modelling a further choice was made to focus on the transport sector, which is the sector with the fastest growth in energy use. With respect to regions, we focus on Western Europe and India, to represent both an industrialised and a developing region.

In this paper, first in Section 2 we discuss the role of uncertainty in energy modelling and introduce a methodology to capture uncertainty in model calibration. In the second part of the article, we elaborate on the application of the method: Section 3 describes the structure of the TIMER 2.0 energy demand model and selects parameters that are

useful for model calibration. Section 4 presents the results of the analysis, Section 5 evaluates the presented methodology and Section 6 discusses and concludes. The underlying details of this paper, including mathematical descriptions, parameter ranges, more in-depth analysis of results and application to the USA, Brazil, China and Russia can be found in [63].

2 Uncertainty in Model Calibration

2.1 Uncertainty in Energy Models

Exploration of different futures on the basis of models is complicated by inherent uncertainties [19, 31, 45–47, 57–60]. Uncertainty and associated terms (such as error, risk and ignorance) are defined and interpreted differently by different authors [for reviews see 28, 46, 56, 68]. These different definitions partly reflect the underlying traditions and their associated scientific philosophical way of thinking. In general, uncertainty may be identified of input parameters, model structure or even different theories at a more aggregated level. Part of these uncertainties are related to natural randomness (ontic). Other uncertainties results from limited knowledge (epidemic). One phase of model development where uncertainties become apparent is during model calibration. Model calibration and validation are of critical importance. As Oreskes et al. [42] highlight, “In areas where public policy and public safety are at stake, the burden is on the modeller to demonstrate the degree of correspondence between the model and the material world it seeks to represent and to delineate the limits of that correspondence.” However, given existing uncertainties in most cases historic trends and data can be interpreted in different ways. This is also emphasised by Beck [4] when he noted that almost all models suffer from a lack of identifiability, i.e. many combinations of values for the model’s parameters may permit the model to fit the observed data more or less equally well.

The notion of ambiguity in model identification and calibration can be valued differently [3, 18]. In statistical modelling traditions, ambiguity in model calibration is typically interpreted as over-parameterisation of the model. Following Occam’s razor, this could be solved with model reduction [11, 26, 71, 72] or developing multiple specialised models [5] to strike a balance between model complexity and data availability. In rule-based (system-dynamic) and engineering models¹ the model

structure is based on (intuitive) causal relations and rules (either in physical or in monetary terms) that are calibrated to historic data [15, 41]. Such causal relations may be postulated, even in the absence of sufficient data for calibration. Beven [7] aims to extend traditional schemes with a more realistic account of uncertainty and rejects the idea that a single optimal model exists for any given case. Instead, models may not be unique in their accuracy of both reproduction of observations and prediction (i.e. unidentifiable or equifinal) and subject to only a conditional confirmation, due to e.g. errors in model structure, calibration of parameters and period of data used for evaluation.

In energy modelling literature, the most analysed sources of uncertainty are parameters and model structure in direct relation with future projections of model drivers. As a typical example, Tschang and Dowlatabadi [53] deal with input parameter uncertainty when performing an uncertainty analysis of the Edmonds–Reilly global energy model. They use Bayesian updating techniques to filter out model simulations that do not conform to outputs on energy consumption and carbon emissions and determine updated prior distributions for several core parameters. Van Vuuren et al. [65] use a slightly more complicated method, in which sampling of input parameters is made conditional upon different consistent descriptions of the future. With respect to model structure, an example is provided by Da Costa [12] who compares the results of two different energy models for Brazil. He concludes that although the aggregate results of these models are comparable, considerable differences exist when the results are broken down.

This study focuses on uncertainty that originates from the calibration of parameter values. We explore whether acceptable sets of parameter values in model calibration (so-called behavioural sets) can be identified for the TIMER energy demand model and what these imply for the model’s projection, inspired by Beven’s work on equifinality.

It should be noted that the mismatch between model prediction and observation can stem from many different sources [7], including those related to measurement, random error, but also the representation of reality by the model as a results of both parameter error and model structure. To keep our analysis manageable, here we assume that the parameter error is the dominant error component—and focus on the question whether our calibration procedure can indeed identify multiple, equally valid, calibrations of the energy demand model. Techniques exist to overcome this simplification and better deconstruct the mismatch between observation and prediction into the six constituting error terms of Beven [7] but this is beyond the scope of the present paper.

¹ Also, especially global energy models are highly policy relevant and are applied for multiple purposes (for instance looking into carbon emission, total energy use, structure of energy use or costs of mitigation measures). This implies that not all model parameters influence the results of all outputs. Hence, these models are de-facto over-parameterised.

2.2 Methodology to Identify Calibrated Sets of Parameter Values

We developed an automated parameter estimation procedure in order to explore the impact of different sets of parameter values on model outcomes. The aim of the developed parameter estimation methodology is two-fold. First, it is an automated model calibration procedure that minimises the error between model results and observations, generating a set of calibrated parameter values. In this sense it is related to nonlinear regression methods like PEST [16] or UCODE [43]. Second, by repeatedly applying the method it can be used to perform an uncertainty analysis on model calibration. This generates a series of calibrated sets of parameter values. This aspect is more related to (sequential) Monte-Carlo-based methods like GLUE [8] or SimLab [49]. The procedure closely follows the manual model calibration process that is normally applied to the TIMER model. This method involves several steps:

- A. Determining useful parameters for model calibration and their associated ranges
- B. Performing a series of model calibrations and identify sets of input parameters that perform well against historic data
- C. Analysing the sets of calibrated parameter values
- D. Analysing the impacts of calibration uncertainty on future projections

2.2.1 Determining Useful Parameters for Model Calibration and Their Associated Ranges

The first step of the method involves analysis of the model, to select useful parameters for the model calibration process. We also identify ranges for the calibration parameters, based on analysis of the model formulation, the values used in former calibrations, literature and expert judgement. This step is described in detail in Section 3. These ranges are used as boundaries in the parameter estimation process.

2.2.2 Performing a Series of Model Calibrations and Identify Sets of Input Parameters that Perform Well Against Historic Data

Criteria for Calibration Fit Several measures exist to evaluate the deviation between model results (predictions, P) and observed data (O), of which an overview can be found in Janssen and Heuberger [27]. We choose to use the normalised root mean square error (NRMSE), comparing

individual time series of observations and predictions, and defined as:

$$\text{NRMSE} = \sqrt{\frac{\sum_{t=1}^T \left(\frac{P_t - O_t}{O_t} \right)^2}{T}} \quad (1)$$

In this, P_t and O_t indicate the predicted and observed value in year t and T is the number of years in the time series. This measure has values between zero (perfect fit) and infinite (random). Multiplied with 100, the NRMSE can be seen as the time averaged percentage deviation between the time series of model results and the time series of observations. A certain threshold level for the NRMSE can be defined, below which models are called *behavioural* with the data (e.g. a NRMSE < 10%), but in [63] we show that it is hardly possible to find criteria for such general numeric threshold.

We use the NRMSE for several reasons. First, it expresses model error at the individual data level. The alternative, expressing model error on the average level, only provides a rough impression of the model-data-discrepancy and averages out the dynamic features [27], whereas with calibration one wants to simulate both trends and patterns in the data. Second, the NRMSE can easily be normalised in each year to observed energy use to prevent that years with higher energy demand dominate the estimated overall error.

Series of Model Calibrations As starting point for the parameter estimations, we use the initial dataset (SI) for P parameters and N parameter estimation attempts: $SI_{P,N}$ (i.e. for the parameter and ranges identified in the previous step). We use a combination of design of experiments (central composite design [39], to explore the extremes of the parameter space) accomplished with a series of random numbers. In the model calibrations, the input parameters are varied in order to minimise the NRMSE, starting at the locations in the parameter space defined in the dataset $SI_{P,N}$. We look for optimal parameter estimations by using a MATLAB built-in functionality for constrained nonlinear optimisation, using sequential quadratic programming [37]. This algorithm approaches the model as a black-box optimisation function and varies the parameter values until the derivative of the objective function (i.e. the NRMSE) reaches values between zero and a pre-defined threshold level. This results in a dataset with calibrated parameter values that have a good (or best obtainable) fit with observations of energy use for the period 1970–2003: $SC_{P,N}$. This can be best imagined as the collection of local optima in the objective function landscape spanned up by the explored parameter space.

2.2.3 Analysis of Calibrated Parameter Values

We analyse the series of calibrated sets of parameter values in $SC_{P,N}$ in several ways. First, the distribution of the calibrated parameter values over their range is analysed. Second, we plot the calibrated parameter values against the NRMSE (see Fig. 3, upper graphs). Relations between parameters and the impact of parameters on the NRMSE can be numerically expressed by the (linear) Pearson correlation coefficient between parameters. We use this as the simplest indicator to express a relation between two parameters, although it does not capture non-linearity or the existence of multimodal distributions.

Based on this, behavioural sets of parameter values can be selected. The most straightforward method is based on the NRMSE value, for instance, one can decide to call sets of parameter values with $NRMSE < 10\%$ behavioural. An alternative, but less reproducible criterion is based on visual inspection of the parameter values and the observed and simulated time series of energy demand. In our analysis, we decided not to remove any sets of parameter values based on non-behavioural outcomes. However, we use the NRMSE (hence, behavioural/non-behavioural) to weight future projections that are derived from the different sets of parameter values.

2.2.4 Analysing the Impacts of Calibration Uncertainty on Future Projections

To analyse the impact of different parameter values on future projections of the model, we use the series of calibrated sets of parameter values in $SC_{P,N}$ to run the model forward for the period 2003–2030 using a similar scenario on the model drivers (see Section 4.2). This leads to a range of projected future energy use, based on the different sets of parameter values. We analyse this in a frequency diagram of energy use in 2030 and weigh the frequencies in the diagram relative to the NRMSE of the parameter set that obtained the best fit to historic data in $SC_{P,N}$ (implicitly assuming that sets of parameter values with a better fit to historic data lead to more plausible future projections). The weight (W) that the N 'th calibrated parameter set gets in the prediction ensemble is defined as the normalisation of the relative weight (R) of the parameter set to the best performing parameter set²:

$$W_N = \frac{R_N}{\sum_N R_N} \text{ where } R_N = \frac{NRMSE_{best}}{NRMSE_N} \quad (2)$$

² This measure does not hold in the unlikely situation that the model exactly reproduces historic data and the best obtained fit becomes zero.

In the remainder of this article, we apply this method to the transport sector energy use model of TIMER 2.0.

3 The TIMER Energy Demand Model

The global energy model TIMER includes both demand and supply of energy [14, 64, 67]. Because of the many feedbacks, interactions and sub-modules, the TIMER model as a whole is too complex to analyse the uncertainty from calibration. Therefore, we here confine the analysis to the sub-model that simulates the demand for energy on the basis of economic activity and autonomous and price-induced efficiency improvements.

Energy use is first modelled as the annual demand for useful energy³ (UE, in GJ/year, see Fig. 1), which is converted to secondary energy use, using specific efficiencies for different fuels. Useful energy demand is modelled as function of four dynamic factors: structural change, autonomous energy efficiency improvement (AEEI), price-induced energy efficiency improvement (PIEEI) and price-based fuel substitution. Thus:

$$UE_{R,S,F(t)} = POP_{R(t)} \cdot X_{R,S(t)} \cdot Y_{R,S,F(t)} \cdot AEEI_{R,S,F(t)} \cdot PIEEI_{R,S,F(t)} \text{ (GJ/yr)} \quad (3)$$

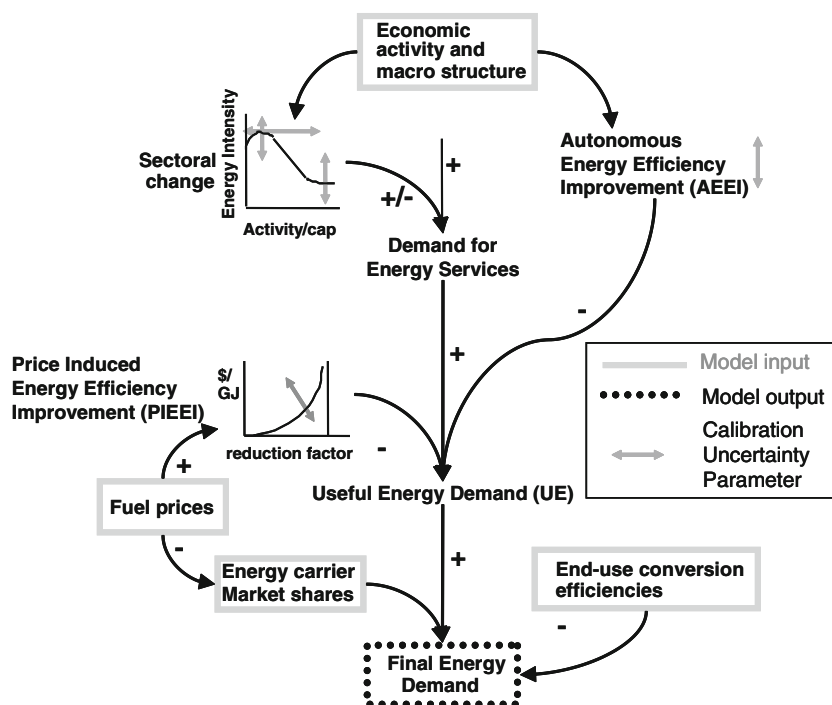
in which POP is the population (in persons), X is the per capita economic activity of a sector (in purchasing power parity (PPP), constant 1995 international \$/capita/year), useful energy intensity (Y , in GJ/\$/capita/year) captures intra-sectoral structural change and the AEEI and PIEEI (dimensionless) multipliers represent autonomous and price-induced efficiency improvements. The indices R , S and F respectively indicate region, sector and energy form (heat or electricity).

Statistical time series are available for several input variables (economic activity, fuel prices and market shares) and for the output variable: final energy use. Between these observed variables, the model tells a story of useful energy intensity (structural change) and autonomous and price-induced efficiency improvements, aggregates that can hardly be measured in the real world. The multiplicative structure of this model leaves room for different behavioural sets of parameter values: for different implementations of UEI, AEEI and PIEEI, a similar result can be obtained for the observed time series of final energy use.

This generic model is used for the five economic sectors in TIMER. In this analysis we look specifically into the

³ With useful energy defined as the level of energy services or energy functions, for instance a heated room or cooled food; conversion efficiencies are taken from statistics.

Fig. 1 Overview of the TIMER energy demand model and identification of model inputs, output and parameters used to determine calibration uncertainty



transport sector implementation of the model. We determine calibration uncertainty against the total demand for transport energy as provided by International Energy Agency (IEA) data. Data for energy prices are derived from the IEA and data on economic activity are obtained from the World Bank WDI [70]. We equate energy demand and energy use, as the statistical data are assumed to have satisfied demand in a state of economic equilibrium on an annual basis; hence, we do not consider the concept of latent (or unfulfilled) demand for energy (which is relevant for low-income regions). Compared to specialised models for transport energy use [e.g. 1, 50, 69], the TIMER model is aggregated and stylized. Especially because it does not take into account the intermediate variables of car ownership or person and freight kilometres or generic concepts like time and money budgets. It also includes energy use for both passengers and freight in a single model.

3.1 Useful Energy Intensity Curve

From energy analysis [see for instance 20, 38, 44] it is known that:

1. there is a tendency for total energy use to increase with population and economic activity
2. in many countries, energy intensity tends first to rise then decline; this takes place at the level of the whole economy but also at sector level. This pattern is often referred to as the Environmental Kuznets Curve [for discussions see 52, 62]. It is usually explained in

structural change processes [For analyses see e.g. 21, 29, 51, 66]. The income level at which such a maximum in intensity is reached tends to decrease over time [6, 22]

These stylized facts are represented in the model equation for useful energy intensity ($Y_{(t)}$) in the form of a (asymmetric) bell-shaped function of the sector-specific per capita economic activity. For each region (R), sector (S) and energy form (F) at time t , this can be expressed as⁴:

$$Y_{R,S,F(t)} = Y_0 + \frac{1}{\beta \cdot X_{(t)} + \gamma \cdot X_{(t)}^\delta} \quad (4)$$

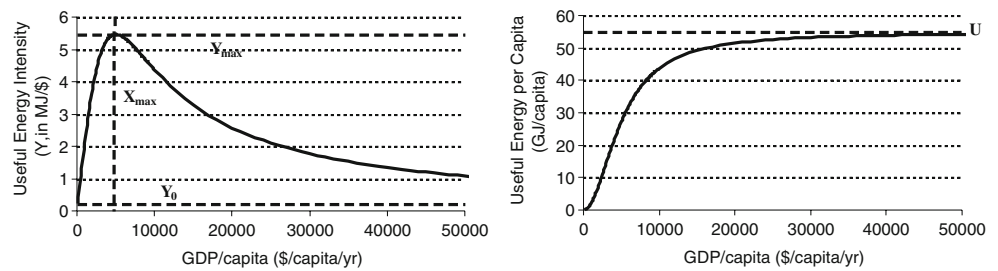
with $X_{(t)}$ the sectoral economic activity per capita and β , γ and δ parameters (of which δ is negative to maintain a bell-shaped form, see Fig. 2). All parameters in this equation are defined per region, sector and energy form.

The flexible formulation of this curve implies also a high sensitivity to parameter values. From an energy point-of-view, some reasonable constraints can be made to limit the potential parameter space to a relevant subspace and to shape the curve on the basis of understandable quantities:

- The activity level at which the maximum occurs, X_{\max} , can be estimated from regional energy use data.

⁴ This bell-shaped curve can also be written in terms of elasticity with GDP/capita as is common for energy use, but for the transport sector it can also be done for vehicle ownership [13].

Fig. 2 UEI curve (left) and useful energy use per capita (right) for hypothetical parameter values



- The second term of the curve may be related to the saturation level of useful energy per capita per year at high income levels (U , see Fig. 2, right graph). This saturation level can be based on sector and region specific features such as climate or population density.
- Y_0 can be interpreted as the ultimately lowest energy intensity of sectoral activity (in \$/GJ) in the both limits $X \rightarrow \infty$ and $X \rightarrow 0$.

Values and ranges for the parameters β , γ and δ can be derived from these constraints, in combination with the assumption that the curve should be forced through one observed reference point which can be any year in the period 1971–2003⁵. Each implementation of the curve (as function of X_{\max} , U and Y_0) can be characterised by its unique maximum energy intensity, i.e. the top of the curve (Y_{\max} , see Fig. 2), determined by:

$$Y_{\max} = Y_0 + \frac{U}{X_{\max}} \cdot \frac{\delta}{(\delta - 1)} \quad (5)$$

We established suitable prior ranges for the variables X_{\max} , U and Y_0 and translated these into values for the curve parameters β , γ and δ [63].

3.2 Autonomous Energy Efficiency Improvement

The continuous decline of energy intensity due to technology change is represented in the TIMER model by the autonomous energy efficiency improvement multiplier. Marginal AEEI is defined as fraction of economic activity growth [34]:

$$AEEImarg_{R,S(t)} = F_S \cdot \left(\frac{GDP_{pc(t)R}}{GDP_{pc(t-1)R}} - 1 \right) \cdot 100 \quad (\%/year) \quad (6)$$

with F_S a sectoral specific fraction of economic activity growth. The vintage structure modelling for energy using capital in TIMER determines that the current AEEI is the weighted average of the marginal AEEI over the capital life time [14]. This means that rapid economic growth leads to a

faster decline in AEEI, due to both increased decline in the marginal AEEI and a larger share of the capital stock that is relatively new [64]. The parameter that can be used to calibrate the AEEI is the fraction of GDP growth (F_S)⁶ (see [63] for details).

3.3 Price-Induced Energy Efficiency Improvement

The PIEEI reflects that with increasing energy prices end-users take measures to use energy more efficiently. The description of PIEEI in TIMER is based on an assumed energy conservation supply–cost curve. By comparing the gains of efficiency improvement (annual saved energy times payback time and energy prices) to the cost of investments, an optimum can be found. There are three main factors that determine the level of energy efficiency: first, the form of the supply–cost curve; second, the value of the payback time and third, learning-by-doing of energy efficiency technology. In the TIMER model, the energy conservation supply–cost curve can be compared to bottom–up technology data [14] but is modelled as an aggregated stylized function. The optimal level of energy efficiency (E , as fraction of total energy use) is defined as the point at which marginal energy conservation measures still yield net revenue:

$$E_{R,S,F(t)} = M_{R,S,F} - \frac{1}{\sqrt{M_{R,S,F}^{-2} + \frac{C_{R,S,F(t)} \cdot T_{R,S,F(t)}}{S_{R,S,F} \cdot I_{R,S,F(t)}}}} \quad (7)$$

in which M is the maximum potential price-induced efficiency improvement (as fraction of total frozen energy use), C the sectoral average costs of useful energy (in \$/GJ) and T the (apparent or desired) payback time (in years). I is the dimensionless factor with which the cost curve declines as a result of learning-by-doing. The scaling parameter S is used to scale the curve to the sector-specific costs of useful energy. The PIEEI on marginal investments, which is used in Eq. 3, is a dimensionless multiplier defined as: $1 - E_{R,S,F}$. Vintage modelling of energy demand capital delays the

⁵ In our model implementation this is the year 2003, the latest year of the calibration period.

⁶ In discussing the results, AEEI is expressed as the average percentage of annual sectoral efficiency improvement, based on the average historic regional GDP per capita growth for the period 1971–2003

impact of the P_{IEEI}, as the current P_{IEEI} is the weighted average of the marginal P_{IEEI} over the capital life time.

In the parameter estimation procedure we vary values of payback time (T) and the learning parameter (I)⁷ using historic energy prices⁸ (see [63] for details).

4 Application to Transport Energy Use

We tested our method to identify multiple behavioural sets of parameter values to the transport sector energy demand sub-model of TIMER. We performed 100 parameter estimation attempts per region (so $N=100$ in $SI_{P,N}$ and $SC_{P,N}$). Section 4.1 discusses the results of calibration to historic data (i.e. step B and C, explained in Section 2.2). Section 4.2 explores the impact of the calibrated sets of parameter values on future projections (step D of the procedure).

4.1 Calibration to Historic Data

Final energy use of the transport sector in Europe shows an increasing trend, with temporary slower growth after 1980 due to oil price increases. Generally, the model simulates transport energy use in Europe quite well with a best NRMSE of 2.8% (Fig. 3). The fluctuations during the 1980s are also well-captured (Fig. 3, lower graph). The calibrated parameter values vary over a wide range and only U , AEEI and P_{IEEI} have clear relations with the NRMSE, although X_{\max} is generally high and Y_0 is low (Fig. 3, upper graphs). About 5% of the sets of parameter values have an NRMSE higher than 10% and can be identified as outliers on the basis of the parameter values. Generally, the parameter values follow two model stories: the best-fitting sets of parameter values have high values for AEEI (>1%/year) and no P_{IEEI}; a second group has low values for AEEI and high P_{IEEI}. The high correlation coefficient between AEEI/P_{IEEI} and NRMSE (Table 1) also shows these different implementations of the parameter values.

In the 1971–2003 period energy use in the transport sector of India has been growing exponentially. This trend is simulated best with a constant useful energy intensity (in

the 1971–2003 GDP/capita range), AEEI of about 1%/year and no P_{IEEI} (Fig. 4). In relation to the NRMSE, a better fit appears with low values for X_{\max} , high U and high AEEI. There are no systematic relations between parameters (Table 1), except between maximum energy intensity (Y_{\max}) and NRMSE (i.e. a lower Y_{\max} leads to a better fit).

Several issues play a role in estimating the model parameters for developing regions. With respect to the UEI curve, these regions have rather narrow absolute GDP per capita ranges between 1971 and 2003 and they are forced to be below the top of the UEI curve (the lower bound of X_{\max} is 5,000 \$/capita/year). Historically, useful energy intensity might have been constant, but it can be questioned whether such implementation of the model is representative outside the range of historically observed economic activity. Another source for the model error in India (and other developing regions) might be that the TIMER model does not capture some important concepts that are relevant for developing countries (e.g. urban/rural differences and unequal income distribution [62]) and ignores the role of specific technologies (e.g. modal split in transport).

4.2 Impact on Future Projections

To determine the influence of the different sets of parameter values on future projections of the model we calculate the projected energy demand in 2030, using scenario inputs of the OECD environmental outlook scenario [OECD-EO, described in detail in 2, 40]. These scenario inputs include projections for GDP, sectoral value added and population. The OECD-EO is a baseline scenario without new policies on economy and environment, in which energy use is based on moderate projections of population and economy. In this analysis we use the same energy prices for all forward calculations; these prices correspond with the default implementation of this scenario⁹.

The TIMER model was used in its original setting within the OECD-EO study to project development of the future energy system, including energy transport demand. These projections can be very different from the current as (1) TIMER modellers have focused in model calibration not only on the performance of a single region but aimed to have similar parameter settings for different regions and (2) have calibrated to the model projections also against the IEA World Energy Outlook.

⁷ Alternative parameters to vary would be the maximum improvement level (M) or the steepness (S). However, M is based on a theoretical maximum efficiency improvement expressed in energy intensity terms. This is a useful parameter to explore, but has more impact on future projections than on historic calibration. The steepness parameter (S) is used to scale the P_{IEEI} curve to the useful energy costs per sector and is therefore not useful to vary.

⁸ We express these two P_{IEEI} related parameters together as the cumulative efficiency improvement in the period 1971–2003.

⁹ Normally energy prices for future projections are calculated endogenously in the TIMER model based on depletion and learning. In this way, different energy demand projections lead to different energy prices, causing different market shares of fuels and other values for end-use-efficiency and P_{IEEI}.

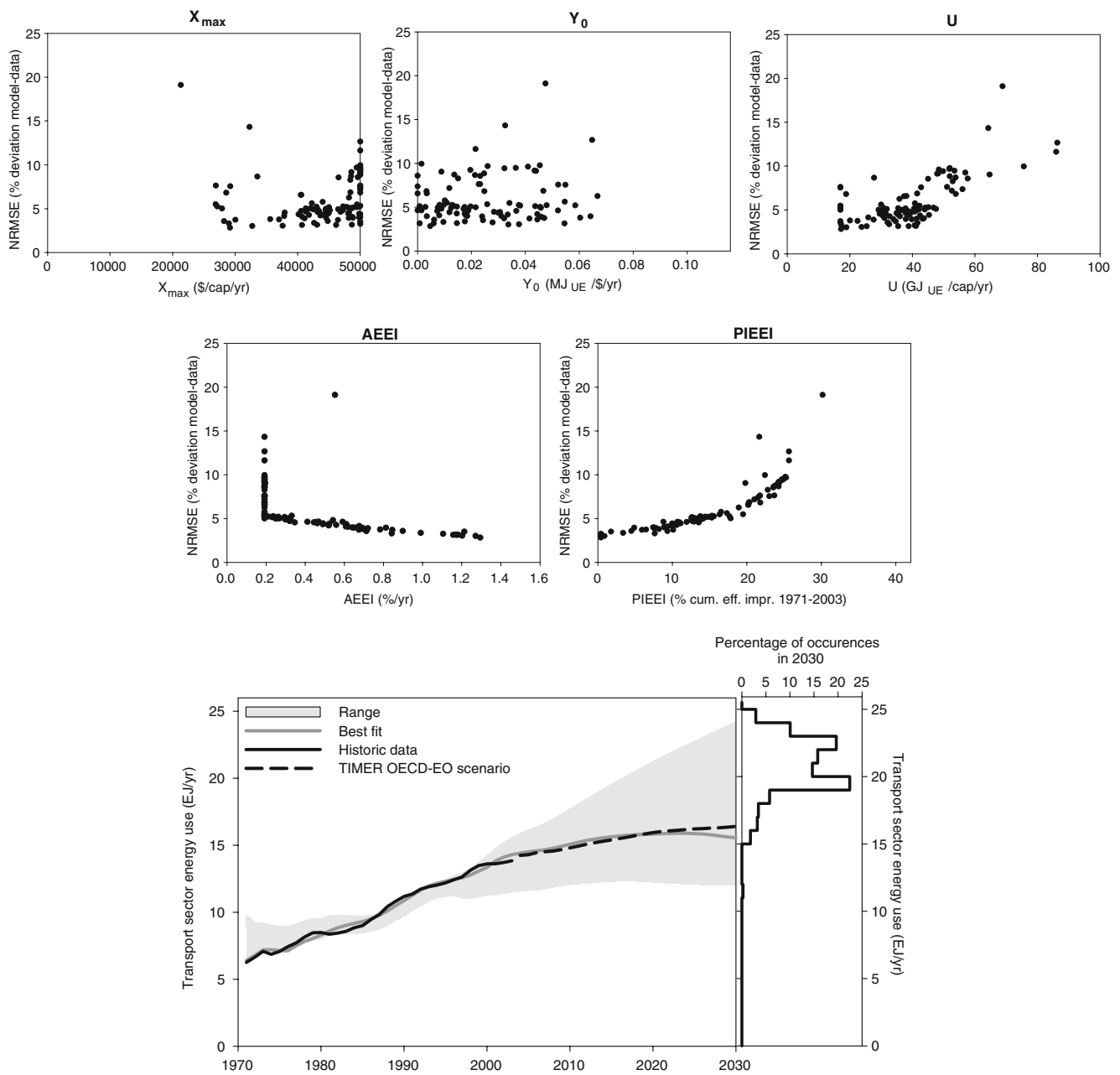


Fig. 3 Upper graphs plot of 100 calibrated sets of parameter values for transport sector energy use in Western Europe. Each dot represents a calibrated parameter value for the period 1971–2003. Lower graphs historic and projected transport energy use for Western Europe up to

2030 (left graph) and histogram (right graph) of energy use in 2030 using the *NRMSE* as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

Table 1 Linear correlation coefficient of calibrated parameter values

Europe	UEI (Y_{\max})	AEEI	PIEEI	India	UEI (Y_{\max})	AEEI	PIEEI
AEEI	0.33	—		AEEI	−0.52	—	
PIEEI	−0.54	−0.86	—	PIEEI	0.06	−0.23	—
<i>NRMSE</i>	−0.52	−0.57	0.85	<i>NRMSE</i>	0.91	−0.75	0.17

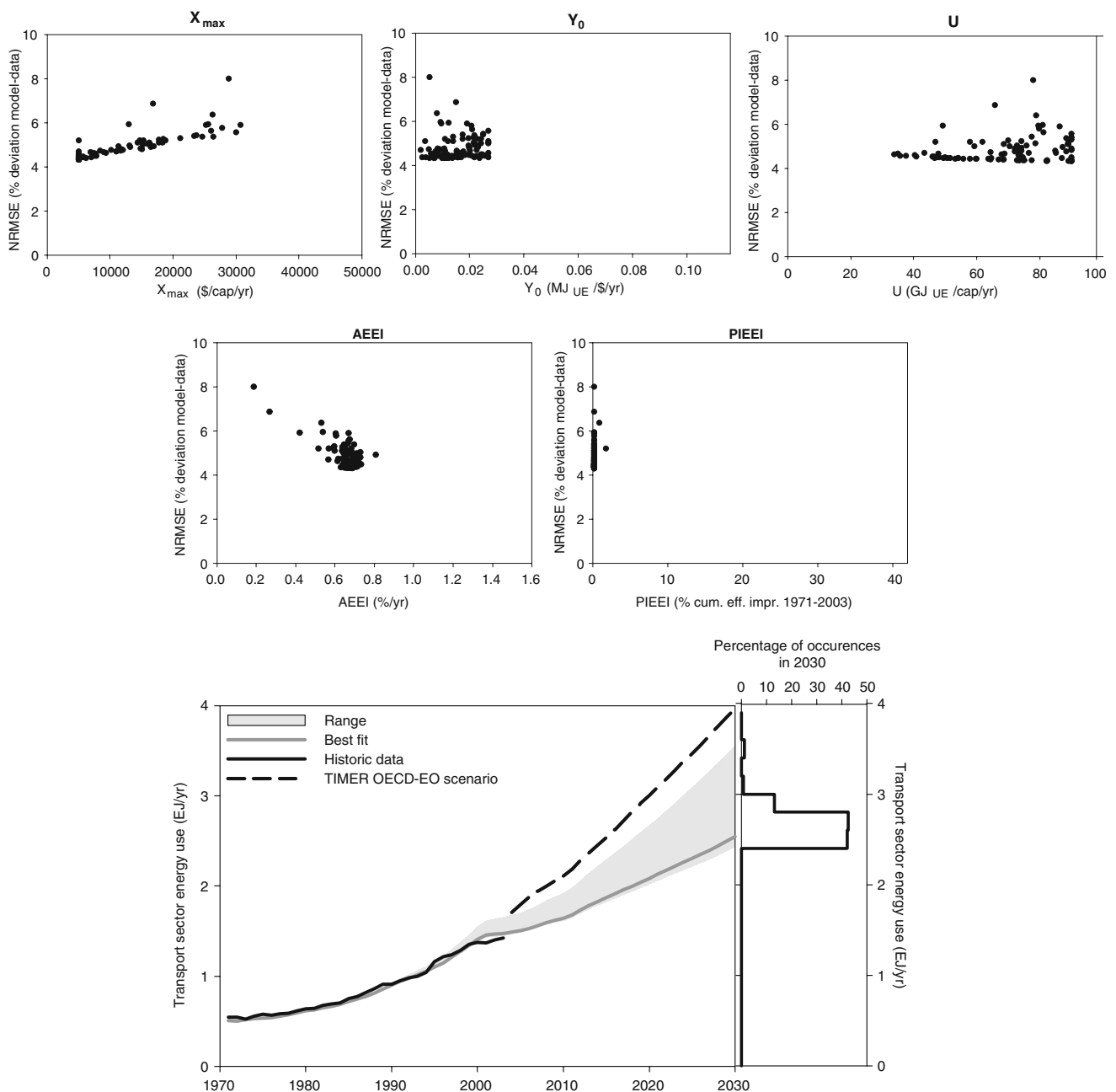


Fig. 4 Upper graphs plot of 100 calibrated sets of parameter values for transport sector energy use in India. Each dot represents a calibrated parameter value for the period 1971–2003. Lower graphs historic and projected transport energy use for India up to 2030 (left

graph) and histogram (right graph) of energy use in 2030 using the NRMSE as weighting factor. Projections based on OECD-EO scenario inputs and calibrated sets of parameter values

The projections of future transport sector energy use in Western Europe in 2030, based on the calibrated sets of parameter values, show a slowly increasing energy use toward 15–25 EJ/year. In 2030, these projections vary over a wide range (Fig. 3); expressed as percentage around the ‘best fit’ in 2030, this range amounts 79% (Table 2). The outliers (with NRMSE values above 8%) are generally on

Table 2 Correlation coefficient between calibrated parameter values and projected energy use in 2030 for the transport sector

	UEI (Y_{\max})	AEEI	PIEEI	Range in 2030
Europe	0.65	−0.32	−0.15	79%
India	0.65	−0.78	0.41	44%

the lower bound of future projections (below 20 EJ/year in 2030). The most behavioural sets of parameter values and the OECD-EO scenario are on the lower bound of this range. However, most sets of parameter values (weighted to the *NRMSE*) project an energy use of 19–23 EJ/year in 2030, higher than the best fitted parameter set and the OECD-EO scenario.

Forward calculations for India indicate an increasing transport sector energy use from 1.5 EJ/year in 2003 to 2.5–3 EJ/year in 2030. Relative to the ‘best fit’, the range for India is narrow: only 44%. The OECD-EO scenario is clearly above the range of projections, leading to 4 EJ/year in 2030. The outliers for India (with *NRMSE* values above 5%) generally project higher future energy use (above 2.8 EJ/year in 2030). Projected energy use correlates strongest with *AEEI* and Y_{\max} : higher *AEEI* (and thus, better fit) leads to lower projected energy use (Table 2).

Another issue of interest is which parameters mainly influence the projected energy use. This is explored in Table 2, showing the correlation between the calibrated parameter values and projected energy use in 2030. *AEEI* and *UEI* seem the most influential model parameters, but the minor role of *PIEEI* might be related to the slow increase of energy prices in the OECD-EO scenario.

5 Method Evaluation

Several remarks can be made about the presented method to identify variation in model calibration parameters. Because the method applies an optimisation algorithm to minimise the error between model results and data, it does not guarantee the identification of the total fit landscape. Especially, if the fit landscape is flat this algorithm identifies the best-fitting (local) optimum, possibly ignoring other well-fitting sets of parameter values that have a slightly higher *NRMSE*. This indicates that the uncertainty from equifinality on forward projections might be larger than estimated in this study. A detailed Monte Carlo sampling analysis would guarantee that the whole fit-landscape is identified. However, we found in early stages of this analysis that equifinality sometimes takes place within very small ranges of the parameter values. Hence, the sampling has to be very detailed in order not to overlook the relevant parameter values, driving up calculation time. We used optimisation to efficiently scan the parameter space, and partly overcome this issue by initialising the parameter estimation process from many different locations in the parameter space (including ‘design of experiments’ to initialise at the corners of the parameter space). However, advanced adaptive sampling methods (see for instance Hendrix and Klepper [24]) might be better able to identify the full range of equifinality.

For this model we chose 100 different initialisations, balancing between calculation time and size of the database. Analysis of the results shows that for this model the shape of the distribution of the parameters and the *NRMSE* did not change significantly after 60 to 80 parameter estimation attempts. We expect this to be specific for each model. If this automated calibration procedure would be applied to another model, convergence of the *NRMSE* and the shape of the parameter distributions should be monitored to see whether enough initialisations have been chosen. It is clear that the method also identifies outliers, cases in which the optimisation algorithm is terminated at relatively high *NRMSE* values. In the analysis that we performed, about 5–10% of the calibrated sets of parameter values could be identified as outliers. We conclude that the estimation technique performs well and most of the identified variation can be attributed to the model at hand.

In the error model that we use, we oversimplified by attributing the difference between modelled and observed values completely to the parameter error. One could extend the method towards more focus on measurement error in the observation (both economic and energy use data are far from certain), for instance by adding white noise to the calibration variable, or input and boundary condition error. In the specific case of *TIMER*, an error distribution on the reference energy intensity for the *UEI* curve might deal with data error and allow a broader range of sets of parameter values to be behavioural with the data. Another issue in the *TIMER* case is that parameter error and model structure error can hardly be separated, because the parameters related to the *UEI* curve, can change the functional form of the model dramatically (e.g. from bell-shaped to linear).

The development of the described method is inspired by the concept of equifinality, developed by Beven based on his experiences with the *GLUE* methodology. The *GLUE* methodology has recently been subject of a scientific debate on its consistency with Bayesian statistics. A major criticism on *GLUE* was its application of ‘less formal likelihood’ measures; this may imply that it loses the learning properties of the Bayesian approach, leading to ‘flat’ parameter posterior densities and thus equifinality is build in the methodology [35, 36]. In response, it has been argued that if strong assumptions about the error model cannot be justified, *GLUE* provides a reasonable alternative [9]. The method applied here differs from both Bayesian updating and *GLUE*, because it does not apply sequential Monte Carlo analysis. Moreover, it also has elements of nonlinear regression methods like *PEST* and *UCODE*, in that its purpose is to identify ‘peaks’ in the fit landscape. Therefore, we conclude that this discussion does not apply to this method.

6 Discussion, Conclusion and Implications

A method was developed to identify sets of parameter values that perform reasonably against historic data. Energy use modelling knows many scientific paradigms and traditions, which lead to different interpretations of past and present and to different expectations of the future. Even within one model, several options may exist on how to interpret the past and current situation. We developed a method to identify the range of sets of parameter values that perform reasonably against historic data and analyse the impact of these different calibrations on future projections. The essence of this method is that by varying several essential parameter values, we search to minimise the error between model results and observations. By repeating this parameter estimation procedure, starting from different locations in the parameter space, we were able to identify a range of local optima in the error landscape within the parameter space. These co-existing different interpretations (i.e. values of essential parameters) that explain historic energy use comparably well are incorporated in the prediction ensemble.

In the energy demand modelling of the TIMER model, different parameters sets can be observed that all lead to reasonable calibration (equifinality). From the application of this method to the TIMER 2.0 energy demand model for the transport sector, we found that its model formulation, in combination with the aggregated character of energy statistics available for calibration, leaves room for multiple behavioural sets of parameter values. In the given model formulation, the different options for calibrated parameter values are related to the balance between useful energy intensity and energy efficiency improvement. Generally, high useful energy intensity combined with major efficiency improvements leads to similar results as low energy intensity and stagnant efficiency improvement.

Different model calibrations lead to different future projections. The range in outcomes is about 44–79% around the best-fit option. With respect to future projections, we found that different (behavioural) sets of parameter values can lead to a wide range of future projections. AEEI and useful energy intensity are the most decisive model aspects with respect to future energy levels.

Equifinality of the TIMER model can partly be improved by further model development. What does this analysis imply for the application and development of the TIMER model? Given the aggregate nature of both model and data some parameter ambiguity is inevitable and does not a priori disqualify the model. For the existing model, a workable situation can be created by using the ‘best-fit’ calibrated parameter values and communicating the calibration uncertainty range with the model results. More fundamentally, two options exist for model improvement. First, the data-based

solution would be model reduction. However, because the model only involves three well-established concepts (energy intensity and autonomous and price-induced efficiency improvement) model reduction implies econometric curve-fitting. A second option is to convert the model to a more bottom-up nature and use the increasingly available data and insights from the underlying physical activity (in this specific case: data on person or freight kilometres, or ownership of cars, trucks, planes etc.; and the concepts of time and money budgets). Such development would lead to two major improvements: first, it provides an extra model layer (of physical activity) that can be calibrated to data and second, such model enhances insight in the actual activity that is simulated and projected.

Acknowledgements This research is financially supported by the Netherlands Environmental Assessment Agency (PBL)

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Azar, C., Lindgren, K., & Andersson, B. A. (2003). Global energy scenarios meeting stringent CO₂ constraints—Cost-effective fuel choices in the transportation sector. *Energy Policy*, 31, 961–976.
2. Bakkes, J., Bosch, P. R., Bouwman, A. F., Eerens, H. E., den Elzen, M., Isaac, M., et al. (2008). Background report to the OECD environmental outlook to 2030. Overviews, details, and methodology of model-based analysis. Bilthoven: Netherlands Environmental Assessment Agency (MNP). 186. <http://www.mnp.nl/bibliotheek/rapporten/500113001.pdf>
3. Barlas, Y. (1989). Multiple tests for validation of system dynamics type of simulation models. *European Journal of Operational Research*, 42(1), 59–87.
4. Beck, B. (2002). Model evaluation and performance. In A. El-Shaarawi & W. Piegorisch (Eds.), *Encyclopedia of environmetrics* (pp. 1275–1279). Chichester: Wiley.
5. Beck, M. B., Ravetz, J. R., Mulkey, L. A., & Barnwell, T. O. (1997). On the problem of model validation for predictive exposure assessments. *Stochastic Hydrology and Hydraulics*, 11(3), 229–254.
6. Bernardini, O., & Galli, R. (1993). Dematerialization: Long-term trends in the intensity of use of materials and energy. *Futures*, 25(4), 431–448.
7. Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1–2), 18–36.
8. Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3), 279–298.
9. Beven, K., Smith, P., & Freer, J. (2007). Comment on "hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini. *Journal of Hydrology*, 338(3–4), 315–318.
10. Bouwman, A. F., Hartman, M. P. M., & Klein Goldewijk, C. G. M. (2006). (Eds.), *Integrated modelling of global environmental change. An overview of IMAGE 2.4*. Bilthoven: Netherlands Environmental Assessment Agency.

11. Crout, N. M. J., Tarsitano, D., & Wood, A. T. (2009). Is my model too complex? Evaluating model formulation using model reduction. *Environmental Modelling & Software*, 24(1), 1–7.
12. da Costa, R. C. (2001). Do model structures affect findings? Two energy consumption and CO₂ emission scenarios for Brazil in 2010. *Energy Policy*, 29(10), 777–785.
13. Dargay, J., Gately, D., & Sommer, M. (2007). Vehicle ownership and income growth, worldwide: 1960–2030. *Energy Journal*, 28(4), 143–170.
14. de Vries, H. J. M., van Vuuren, D. P., den Elzen, M. G. J., & Janssen, M. A. (2001). The TIMER IMage Energy Regional (TIMER) model. Bilthoven: National Institute for Public Health and the Environment (RIVM). 188. <http://www.mnp.nl/bibliotheek/rapporten/461502024.pdf>.
15. Dogan, G. (2004). Confidence interval estimation in system dynamics models: Bootstrapping vs. likelihood ration method. 22nd International Conference of the System Dynamics Society. Oxford, UK.
16. Doherty, J. (2004). PEST model-independent parameter estimation, user manual: 5th edition. Brisbane, Australia: Watermark Numerical Computing. 336. www.sspa.com/pest
17. Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B. Methodological*, 57(1), 45–97.
18. Edwards, P. N. (1999). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science As Culture*, 8(4), 437–472.
19. Filar, J. A. (2002). *Mathematical models. knowledge for sustainable development—An insight into the encyclopedia of life support systems (pp. 339–354). Released at the world summit on sustainable development*. Johannesburg: UNESCO/EOLSS.
20. Focacci, A. (2005). Emperical analysis of the environmental and energy policies in some developing countries using widely employed macroeconomic indicators: The cases of Brazil, China and India. *Energy Policy*, 33, 543–554.
21. Gales, B., Kander, A., Malanima, P., & Rubio, M. (2007). North versus South: Energy transition and energy intensity in Europe over 200 years. *European Review of Economic History*, 11(2), 219–253.
22. Groenenberg, H. (2002). Development and convergence, a bottom-up analysis for the differentiation of future commitments under the climate convention. Faculty of Chemistry. PhD Thesis, Utrecht: Universiteit Utrecht.
23. Grubb, M., Edmonds, J., Brink, P. T., & Morrison, M. (1993). The costs of limiting fossil-fuel CO₂ emissions: A survey and analysis. *Annual Review of Energy and the Environment*, 18(1), 397.
24. Hendrix, E. M. T., & Klepper, O. (2000). On uniform covering, adaptive random search and raspberries. *Journal of Global Optimization*, 18(2), 143–163.
25. IPCC. (2000). *Special report on emission scenarios*. Cambridge: Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
26. Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21(5), 602–614.
27. Janssen, P. H. M., & Heuberger, P. S. C. (1995). Calibration of process-oriented models. *Ecological Modelling*, 83(1–2), 55–66.
28. Janssen, P. H. M., Petersen, A. C., Van der Sluijs, J. P., Risbey, J., & Ravetz, J. R. (2005). A guidance for assessing and communicating uncertainties. *Water Science and Technology*, 52(6), 125–131.
29. Kander, A., & Schon, L. (2007). The energy–capital relation—Sweden 1870–2000. *Structural Change and Economic Dynamics*, 18(3), 291–305.
30. Kann, A., & Weyant, J. (2000). Approaches for performing uncertainty analysis in large-scale energy/economic policy models. *Environmental Modeling & Assessment*, 5(1), 29–46.
31. Kleindorfer, G. B., O'Neill, L., & Ganeshan, R. (1998). Validation in simulation: Various positions in the philosophy of science. *Management Science*, 44(8), 1087–1099.
32. Löschel, A. (2002). Technological change in economic models of environmental policy: A survey. *Ecological Economics*, 43(2–3), 105–126.
33. MA. (2005). *Millenium ecosystem assessment: Ecosystems for human wellbeing*. Washington DC: Island Press.
34. Manne, A., Richels, R., & Edmonds, J. (2005). Market exchange rates or purchasing power parity: Does the choice make a difference to the climate debate? *Climatic Change*, 71(1), 1–8.
35. Mantovan, P., & Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, 330(1–2), 368–381.
36. Mantovan, P., Todini, E., & Martina, M. L. V. (2007). Reply to comment by Keith Beven, Paul Smith and Jim Freer on "hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology". *Journal of Hydrology*, 338(3–4), 319–324.
37. Mathworks (2007). Optimization toolbox, user's guide. Natick, MA, USA. http://www.mathworks.com/access/helpdesk/help/pdf_doc/optim/optim_tb.pdf.
38. Medlock, K. B., III, & Soligo, R. (2001). Economic development and end-use energy demand. *Energy Journal*, 22(2), 77.
39. NIST/SEMATECH. e-Handbook of Statistical Methods. 2006 [cited 2007 5 October]; Available from: <http://www.itl.nist.gov/div898/handbook/>.
40. OECD. (2008). *OECD environmental outlook to 2030*. Paris: OECD. www.oecd.org/environment/outlookto2030.
41. Oliva, R. (2003). Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151(3), 552–568.
42. Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263, 641–646.
43. Poeter, E. P., Hill, M. C., Banta, E. R., Mehl, S., & Christensen, S. (2005). UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. U.S. Geological Survey Techniques and Methods: U.S. Geological Survey.
44. Reddy, A. K. N., & Goldemberg, J. (1990). Energy for the developing world. *Scientific American*, 263(3), 111.
45. Refsgaard, J. C., van der Sluijs, J. P., Brown, J., & van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29, 1586–1597.
46. Refsgaard, J. C., van der Sluijs, J. P., Hojberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process—A framework and guidance. *Environmental Modelling & Software*, 22(11), 1543–1556.
47. Risbey, J., Van der Sluijs, J. P., Klopogge, P., Ravetz, J., Funtowicz, S., & Corral Quintana, S. (2005). Application of a checklist for quality assistance in environmental modelling to an energy model. *Environmental Modeling & Assessment*, 10(1), 63–79.
48. Rotmans, J., & de Vries, H. J. M. (1997). *Perspectives on global change, the TARGETS approach*. Cambridge: Cambridge University Press.
49. Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice, a guide to assessing scientific models*. Chichester: Wiley.
50. Schafer, A., & Victor, D. G. (2000). The future mobility of the world population. *Transportation Research Part A: Policy and Practice*, 34(3), 171–205.
51. Seppälä, T., Haukioja, T., & Kaivo-oja, J. (2001). The EKC hypothesis does not hold for direct material flows: Environmental

- Kuznets Curve hypothesis tests for direct material flows in five industrial countries. *Population and Environment*, 23(2), 217–238.
52. Stern, D. I. (2004). The rise and fall of the Environmental Kuznets Curve. *World Development*, 32(8), 1419–1439.
 53. Tschang, F. T., & Dowlatabadi, H. (1995). A Bayesian technique for refining the uncertainty in global energy model forecasts. *International Journal of Forecasting*, 11(1), 43–61.
 54. UNEP. (2007). *Global environment outlook: Environment for development*. Nairobi: United Nations Environment Program. <http://www.unep.org/geo/geo4>.
 55. van den Berg, H. (1994). *Calibration & evaluation of a global energy model (submodel of TARGETS)*. Centre for energy and environmental studies (IVEM). Groningen: University of Groningen.
 56. van der Sluijs, J. P. (1997). Anchoring amid uncertainty, on the management of uncertainties in risk assessment of anthropogenic climate change. Department of Science, Technology and Society. PhD thesis, Utrecht: Utrecht University.
 57. van der Sluijs, J. P. (2002). A way out of the credibility crisis of models used in integrated environmental assessment. *Futures*, 34(2), 133–146.
 58. van der Sluijs, J. P. (2005). Uncertainty as a monster in the science policy interface: Four coping strategies. *Water Science and Technology*, 52(6), 87–92.
 59. van der Sluijs, J. P. (2006). Uncertainty, assumptions, and value commitments in the knowledge-base of complex environmental problems. In Â. G. Pereira, S. G. Vaz & S. Tognetti (Eds.), *Interfaces between science and society* (pp. 67–84). New York: Green Leaf Publishing.
 60. van der Sluijs, J. P. (2007). Uncertainty and precaution in environmental management: Insights from the UPEM conference. *Environmental Modelling & Software*, 22(5), 590–598.
 61. van der Sluijs, J. P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., et al. (2001). Uncertainty assessment of the IMAGE-TIMER B1 CO₂ emissions scenario, using the NUSAP method: Dutch National Research Program on Climate Change. 225.
 62. van Ruijven, B., Urban, F., Benders, R. M. J., Moll, H. C., van der Sluijs, J. P., de Vries, B., et al. (2008). Modeling energy and development: An evaluation of models and concepts. *World Development*, 36(12), 2801–2821.
 63. van Ruijven, B. J., van der Sluijs, J. P., van Vuuren, D. P., Janssen, P. H. M., Heuberger, P. S. C., & de Vries H. J. M. (2009). Uncertainty from model calibration: Applying a new method to calibrate energy demand for transport. Utrecht/Bilthoven: Utrecht University, Dept. of STS / Netherlands Environmental Assessment Agency (PBL). 33. http://www.chem.uu.nl/nws/www/research/risk/Ruijven_Model_Calibration_Uncertainty.pdf
 64. van Vuuren, D. P. (2007). *Energy systems and climate policy*. Dept. of Science, Technology and Society, Faculty of Science. Utrecht: Utrecht University.
 65. van Vuuren, D. P., de Vries, B., Beusen, A., & Heuberger, P. S. C. (2008). Conditional probabilistic estimates of 21st century greenhouse gas emissions based on the storylines of the IPCC-SRES scenarios. *Global Environmental Change*, 18(4), 635–654.
 66. van Vuuren, D. P., Strengers, B. J., & De Vries, H. J. M. (1999). Long-term perspectives on world metal use—A system-dynamics model. *Resources Policy*, 25(4), 239–255.
 67. van Vuuren, D. P., van Ruijven, B. J., Hoogwijk, M. M., Isaac, M., & de Vries, H. J. M. (2006). TIMER 2.0, Model description and application. In A. F. Bouwman, M. P. M. Hartman & C. G. M. Klein Goldewijk (Eds.), *Integrated modelling of global environmental change. An overview of IMAGE 2.4*. Bilthoven: Netherlands Environmental Assessment Agency (MNP).
 68. Walker, W. E., Harremoës, P., Rotmans, J., Van der Sluijs, J. P., Van Asselt, M. B. A., Janssen, P., et al. (2003). Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1), 5–17.
 69. Wohlgemuth, N. (1997). World transport energy demand modelling: Methodology and elasticities. *Energy Policy*, 25(14–15), 1109–1119.
 70. World Bank (2004). World development indicators (CD-ROM). World Bank.
 71. Young, P. (1998). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, 13(2), 105–122.
 72. Young, P. C., Parkinson, S. D., & Lees, M. (1996). Simplicity out of complexity: Occam's razor revisited. *Journal of Applied Statistics*, 23, 165–210.